

SAIC0008-CON1
Serial No. 10,760,472

2

PATENT

In the Claims

33. (Previously presented) A system for automatically cataloguing documents located in multiple heterogeneous repositories, the system comprising:

a scanning tool for scanning the multiple heterogeneous repositories to collect keywords for the documents located therein;

a keyword index to the documents built using the collected keywords;

a mapping tool for mapping the documents using the keyword index to one or more classes, each of the one or more classes including keywords representative of that class; and

a computing device for creating metadata indicative of each of the documents and cataloguing each of the documents in an integrated library according to the metadata in a meta-index, wherein the metadata for each of the documents indexed within the meta-index is stored in a pre-defined data structure including at least one of the following attributes a uniform resource locator (URL), a title, an author, an abstract, a collection, a keyword, one or more matched words, a path, a classmark, a classification date and a last modified date.

34. (Previously presented) The system according to claim 33, wherein the meta-index retains characteristics of each of the multiple heterogeneous repositories as applied to each of the documents such that a user may access one or more of the documents within the multiple heterogeneous repositories utilizing the meta-index.

35. (Previously presented) The system according to claim 34, wherein the characteristics of the multiple heterogeneous repositories are transparent to the user when one or more of the documents are accessed using the meta-index.

<US1900> 9148967.1

SAIC0008-CON1
Serial No. 10,760,472

3

PATENT

36. (Previously presented) The system according to claim 33, wherein the metadata is stored in eXtensible Markup Language (XML) format.

37. (Previously presented) The system according to claim 33, wherein the metadata is stored in Resource Description Framework (RDF) format.

38. (Previously presented) The system according to claim 33, wherein the scanning tool is at least one spider.

39. (Previously presented) The system according to claim 33, wherein the mapping tool is a domain ontology.

40. (Previously presented) The system according to claim 39, wherein the domain ontology is a classification hierarchy.

41. (Previously presented) The system according to claim 33, wherein the mapping tool is a neural network.

42. (Previously presented) A method for automatically cataloguing documents located in multiple heterogeneous repositories, comprising:

scanning the multiple heterogeneous repositories to collect keywords from the documents located therein;

<US1900> 9148967.1

SAIC0008-CON1

4

PATENT

Serial No. 10/760,472

building a keyword index to the documents stored in the multiple heterogeneous repositories using the collected keywords;

mapping the documents using the keyword index into predetermined classes, wherein the mapping is performed using at least one mapping tool;

creating metadata information, including identification of the predetermined class, for the documents; and

cataloguing each of the documents in an integrated library according to the metadata in a meta-index, wherein the metadata for each of the documents indexed within the meta-index is stored in a pre-defined data structure including at least one of the following attributes a universal resource locator, a title, an author, an abstract, a collection, a keyword, one or more matched words, a path, a classmark, a classification date and a last modified date and further wherein the meta-index retains the characteristics of each of the multiple heterogeneous repositories as applied to each of the documents such that a user may access one or more of the documents within the multiple heterogeneous repositories utilizing the meta-index.

43. (Previously presented) The method of claim 42, wherein scanning the at least one information repository to collect keywords is performed by a spider.

44. (Previously presented) The method of claim 42, wherein the metadata information is stored in the eXtensible Markup Language (XML) format.

45. (Previously presented) The method of claim 42, wherein the metadata information is stored in the Resource Description Framework (RDF) format.

<US1900> 9148967 2

SAIC0008-CON1
Serial No. 10,760,472

5

PATENT

46. (Previously presented) A method for automatically cataloguing documents located on at least a first and second website, comprising:

scanning the at least a first and second website to collect keywords from the documents located therein, wherein documents located on a first website are in a first format and documents located on a second website are in a second format;

building a keyword index to the documents stored on the at least a first and second website using the collected keywords;

mapping the documents using the keyword index into predetermined classes, wherein the mapping is performed using at least one mapping tool;

creating metadata information, including identification of the predetermined class, for the documents; and

cataloguing each of the documents in an integrated library according to the metadata in a meta-index, wherein the metadata for each of the documents indexed within the meta-index is stored in a third format and further wherein the meta-index retains the first format and the second format, respectively, for the documents in each of the at least a first and second websites such that a user may access one or more of the documents within the at least a first and second website utilizing the meta-index.

47. (Previously presented) The method of claim 46, wherein scanning the at least a first and second website to collect keywords is performed by a spider.

<US1900> 9148967.1

SAIC0008-CON1**6****PATENT**

Serial No. 10,760,472

48. (Previously presented) The method of claim 46, wherein the metadata is stored in the eXtensible Markup Language (XML) format.

49. (Previously presented) The method of claim 46, wherein the metadata is stored in the Resource Description Framework (RDF) format.

<US1900> 9148967.1